

High-throughput purification of hexahistidine-tagged proteins expressed in *E. coli*

Michael B. Murphy and Sharon A. Doyle

Running title: High-throughput Protein Purification

Keywords: IMAC, protein production, hexahistidine tag

Sharon A. Doyle, Ph.D.
DOE Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
Tel: 925-296-5694
Fax: 925-296-5850
Email: sadoyle@lbl.gov

1. Introduction

Large collections of purified proteins have become essential to systems biology programs for generating protein ligands and for validating protein interactions. These large protein collections will likewise benefit structural studies, and quests for protein-based therapeutics. Whether many protein targets need to be produced, such as a microbial proteome, or a few poorly expressing protein targets need to be expressed as soluble fragments (divide and conquer approach), efficient high-throughput processing can be a bottleneck (1).

This chapter describes a method for efficient high-throughput purification of hexahistidine-tagged proteins that are expressed in *Escherichia coli* (*E. coli*) using immobilized metal affinity chromatography (IMAC) (2) in a 96-well format. This approach is particularly suitable for proteomic applications that require modest amounts of highly purified proteins to be generated very efficiently. This approach is also very useful for identifying protein targets that are most amenable to scaled-up production for use in structural studies. The typical yield of proteins purified using this system is 50–150 micrograms, which is generally greater than that of many *in vitro* expression systems and much less costly. The method as described has been optimized for purifying approximately 150 micrograms of hexahistidine-tagged protein, but the method is flexible so that the amount of affinity matrix and culture volumes can be adjusted for optimal binding capacity and consequently highest purity. Although the method detailed here uses IMAC to purify hexahistidine-tagged proteins, this basic platform can be used with many other tags and affinity resins.

2. Materials

1. 24-well (7 ml) and/or 96-well (2 ml) round bottom blocks (Qiagen, Valencia, CA)
2. Genemate 96-well filter-plate with GMF-5 filter (ISC Bioexpress, Kaysville, UT)
3. AirPore Tape (Qiagen, Valencia, CA)
4. 96-well plate shaker/vortex
5. 96-well plate/block centrifuge
6. 96-well filter-plate vacuum manifold and regulator (Qiagen, Valencia, CA)
7. LB media and LB agar plates
8. Kanamycin
9. Isopropyl-beta-D-thiogalactopyranoside (IPTG)
10. Lysis Buffer: 50 mM sodium phosphate buffer (NaH_2PO_4), pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.05% polysorbate 20, 1 mM MgCl. Store at 4 °C.
11. Wash Buffer: 50 mM NaH_2PO_4 , pH 8.0, 500 mM NaCl, 20 mM imidazole, 0.05% polysorbate 20. Store at 4 °C.
12. Elution Buffer: 50 mM NaH_2PO_4 , pH 8.0, 150 mM NaCl, 500 mM imidazole, 10% glycerol. Store at 4 °C.
13. Lysozyme (Sigma, St. Louis, MO)
14. Benzonase (Novagen, Madison, WI) or DNase
15. Protease inhibitor cocktail (for histidine-tagged proteins) (Sigma-Aldrich, St. Louis, MO)
16. Phenylmethylsulfonylfluoride (PMSF)
17. Ni-NTA Superflow resin (Qiagen, Valencia, CA) or other IMAC resin

18. BL21 Gold (DE3) competent cells (Stratagene, La Jolla, CA)
19. 3 –15 % polyacrylamide gels (BioRad, Hercules, CA)
20. GelCode Blue (Pierce, Rockford, IL)
21. Agilent 2100 Bioanalyzer and Protein 200 plus lab-chip kit (Agilent, Palo Alto, CA)

3. Methods

First, a general method of transformation of the expression plasmid into *E. coli* will be described as well as appropriate conditions for growth and induction of the expression cultures. Then, the 96-well protein purification method is detailed. Last, analysis of the purified proteins is described using both “lab-on-a-chip” technology and traditional SDS-PAGE.

3.1 Transformation and Expression Induction

Expression plasmids, in our example pNHIS (3), containing genes of interest fused to a hexahistidine coding tag are first purified from 5 ml cultures of a non-expression strain of *E. coli*, such as DH5 α or Novablue (a K-12 strain from Novagen), grown in LB containing kanamycin for antibiotic selection. An appropriate *E. coli* strain for protein expression is then transformed with the purified expression plasmids, as described below. The *E. coli* cell strain that is most commonly used for protein expression driven by T7 RNA polymerase (4) is BL21(DE3). This strain contains a chromosomal copy of the T7 RNA polymerase gene (indicated by the DE3), the expression of which is inducible by addition of IPTG. There are many derivative and mutant *E. coli* strains available that offer strategies for producing problematic proteins (see **Note 1**).

1. Add 1 μ l of purified plasmid DNA to 25 μ l of BL21 (DE3) competent cells in a 1.5 ml microfuge tube, gently mix and place on ice for 10 minutes.
2. Place in a 42 °C water bath for 30 seconds, then immediately back on ice for 2 minutes. Add 500 μ l of SOC (or LB) and grow with vigorous shaking at 37 °C for 30 minutes.
3. Plate 50 μ l onto prewarmed LB agar plates containing 50 μ g/ml kanamycin, and grow overnight at 37 °C.
4. Next, grow starter cultures by picking a colony from the above plates into 2 ml LB plus 50 μ g/ml kanamycin in sterile 24-well blocks. Cover the blocks with airpore tape and grow overnight at 37 °C, shaking at 180-200 rpm.
5. Aliquot 5 ml of LB (*see Note 2*) plus 50 μ g/ml kanamycin into each of the wells of new 24-well blocks. Inoculate the 5 ml of LB with 100 μ l of the starter cultures (*see Note 3*). Cover the blocks with air-pore tape and grow at 37 °C at 180-200 rpm for approximately 3 hours until the optical density (O.D.) at 600 nm wavelength is 0.6-0.8. The OD₆₀₀ should be checked every hour by removing 100 μ l from several representative wells.
6. Add IPTG to induce protein expression, transfer the blocks to a shaker incubator that has been adjusted to the desired expression temperature, and grow for the desired amount of time. Each of these parameters should be optimized by the expression screening technique described in the previous chapter (*see Note 4*). In the example in this chapter, protein expression was induced with 1 mM IPTG and cultures were grown for 6 hours at 25 °C.
7. Pellet the cells by centrifugation at 2500 x g for 10 minutes at 4 °C and then carefully decant all media. Freeze the bacterial pellets at -80 °C.

3.2 Protein Purification

Purification of the proteins from the cultures grown in 24-well blocks can be performed in 96-well filter plates using a vacuum manifold for column forming, washing, and elution steps, as outlined below. The hexahistidine tagged proteins in cleared cell lysates are batch loaded onto Ni-NTA agarose in a 96-well filter plate. Columns are formed by applying low (200 mbar) vacuum pressure. These mini-columns are extensively washed and subsequently the proteins are eluted into a microtiter plate.

1. Thaw bacterial pellets in 24-well blocks at room temperature. While waiting, add protease inhibitors, lysozyme, and Benzonase to the chilled Lysis buffer: 1 mg/ml Lysozyme, 10 units Benzonase/ml, 2 μ l/ml protease inhibitor cocktail (no EDTA), and 1 mM PMSF. Keep on ice.
2. Add 0.5 ml of chilled Lysis buffer plus additives to each pellet and resuspend by pipetting, using a 12-channel pipet to mix a row of 6 wells with two tips in each of the 6 wells.
3. Vortex the 24-well blocks containing the resuspended pellets at 4 °C on a plate shaker for 20 minutes. Performing the lysis at 4 °C will help prevent protein degradation.
4. Centrifuge 24-well blocks at 2500 x g in a plate centrifuge for 20 minutes at 4 °C. The pellets obtained can be saved to purify insoluble protein under denaturing conditions (not discussed here).

5. Equilibrate the Ni-NTA resin in Lysis buffer and resuspend to obtain a 50% slurry. Add 50 μ l of the 50% slurry of Ni-NTA resin, which is enough to bind approximately 150 μ g of protein, into the wells of a 96-well filter-plate (*see Note 5*). Liquid may begin to flow by gravity, in order to prevent loss of material place the filter plate on a piece of parafilm.
6. Carefully transfer the cleared bacterial lysate supernatants from each of the 24-wells into the wells of 96-well filter-plate. Cover the filter plate top with clear plate seal tape, to prevent cross-contamination.
7. Mix the Ni-NTA resin with the cleared lysate in the 96-well filter-plate on a piece of parafilm on a plate vortex for 30 minutes at 4 °C.
8. Remove the 96-well filter-plate from the plate vortex leaving the parafilm behind, and place on the vacuum manifold (with a waste collection container below) and apply vacuum (approximately 200 mbar) so that each well slowly drips over several minutes until all the lysate has filtered through, and release the vacuum after the material has flowed through.
9. Add 750 μ l of Wash buffer (*see Note 6*) into each well and apply the vacuum just as in step 8. Repeat the wash step 2 more times.
10. Fit a 96-well microtiter plate (200 μ l) or 2 ml collection microtubes into vacuum manifold carefully aligning the drop-formers to the receiver plate wells. Add 100 μ l Elution buffer and incubate for 2 minutes. Then apply a vacuum of 200-400 mbar and release vacuum after material has flowed through. Store the plate of eluted proteins at 4 °C for short term or until evaluated and then at –80 °C for long term storage.

3.3 Determination of Protein Yield and Purity

Eluted material was assessed on an Agilent 2100 Bioanalyzer using a Protein 200 Plus LabChip (Caliper) to determine protein purity, concentration, and yield. However, traditional SDS-PAGE can also be used, if necessary. The LabChip system is a fluorescence-based detection system offers the advantages of requiring only 1-4 μ l of protein samples, as compared to 10 μ l required for equivalent band identification on traditional SDS-PAGE gels, and results are obtained in 30 minutes a set of 10 samples. The Biosizing software provides detailed tables of raw and analyzed data and gel-like images, and is convenient for data storage and retrieval. The Agilent 2100 Bioanalyzer system provides accurate protein concentrations, alleviating the need of running additional assays and using more of the samples. In addition, the concentration of each individual protein is determined from the fluorescent peak area, or “bands” on the gel-like image, thus accurate concentrations of a protein of interest can be easily obtained in partially purified samples.

1. For SDS-PAGE, 10 μ l of each protein sample were run on 3-15% gradient polyacrylamide gel. The gels were stained with GelCode Blue and scanned on Fluor-S MultiImager (**Fig 1B**) (BioRad, Hercules, CA).
2. For the Bioanalyzer, the protein LabChip is prepared according to the manufacturer's instructions.
3. The samples were prepared by mixing 2 μ l of eluted protein, 2 μ l H₂O and 2 μ l of a SDS-based denaturing sample buffer containing β -mercaptoethanol as well as upper and lower protein mass standards. Samples are heated to 95 °C for 2 minutes and spun briefly. (*see Note 7*).

4. Samples and ladder are then diluted to 90 μ l with H₂O and then 6 μ l of each diluted sample is loaded into a well of the LabChip according to the manufacturers instructions. Dilution of the sample may seem counter-intuitive but is necessary to reduce the SDS concentration in the denaturation buffer.
5. The LabChip is then placed in an Agilent 2100 Bioanalyzer and the Protein 200 Plus program is run.
6. Agilent Biosizing software is used to determine the sizes of the proteins of interest by normalization against two internal standards of 6 and 210 kDa. The fluorescent peak identification settings were adjusted from the default settings to improve sensitivity, 0.8 for the minimum peak height, 0.2 seconds for the minimum peak width, and 4 for the slope threshold. (**Fig. 1A**)

4. Notes

1. A variety of *E. coli* strains have been developed for protein expression that may improve the soluble yields of particular classes of proteins. We have found that many microbial proteins express well in the traditional BL21(DE3) strain. However, eukaryotic proteins often express better in strains that provide tRNAs for codons that *E. coli* rarely use, such as Rosetta (Novagen). Other examples include mutant strains that may improve the proper folding of some proteins by promoting disulfide bond formation, such as Origami and Origami B (Novagen), and mutant strains that facilitate IPTG concentration dependent expression induction, such as Tuner (Novagen). Testing several expression strains from various vendors is often a good strategy to improve the odds of producing

soluble protein and further highlights the utility of high-throughput expression and purification techniques.

2. A variety of uniquely formulated media are available that may improve growth. For example, richer media may be used such as Power, Turbo or Superior broth (Athena ES, Baltimore, MD) to increase biomass and subsequent protein yield.
3. Replicates of each of the proteins can be set-up in the 24-well blocks and then combined into the same well of a 96-well block to increase the yield of low expressing samples.
4. The chapter on expression screening describes a simple method to quickly identify the optimal condition for expression induction. The amount of IPTG (0.1 to 1 mM), optimal temperature (18 °C to 37 °C), and length of the induction time (4 hr to overnight) should be identified to improve the likelihood of producing soluble protein and improve the yields of soluble proteins. Additionally, purification from several protein expression conditions can be tested in parallel.
5. IMAC resins are available from many sources, and the binding capacity can vary. Additionally, use of metal chelation via tetradentate adsorbant (NTA) as opposed to tridentate (IDA) ensures that much less leaching of metal ions will occur. The purity of hexahistidine tagged proteins is optimal from purification conditions in which the amount of the target protein is equal to or greater than the binding capacity of the column matrix. The amount of lysate and Ni-NTA Superflow agarose that we have suggested should be a guide. The expression screening method described in the previous chapter can be used to assess and then group samples with similar expression levels and then approximate the optimal amount of Ni-NTA Superflow agarose required.

Additionally, we tested several 96-well filter plates and found that the Genemate plates have dropformers that perform better than others with less cross-contamination and optimal recovery of eluent, and are less expensive than most other plates.

6. The compositions of the buffers can be varied. Adjusting the stringency of the wash, for example by increasing the NaCl concentration to 500 mM or the imidazole concentration to 20 mM can improve purity, however, lowering these concentrations may improve binding of some hexahistidine-tagged proteins. The QIAexpressionist handbook (Qiagen, Valencia, CA) provides some very useful suggestions for various buffer compositions.
7. The LabChip protocol recommends using 4 μ l of protein sample, however we find that using 1-2 μ l of each sample plus H₂O works as well or better, and uses less material. This may be due to a dilution of components that affect the sensitivity of the assay, such as salt and imidazole. Many common reagents affect the sensitivity of the LabChip assay, therefore care should be taken when the buffer composition is altered.

Acknowledgements

We thank Shirin Fuller for technical assistance, and Paul Richardson and Peter Beernink for helpful suggestions and critical reading of the manuscript. This work was performed under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research, by the University of California, Lawrence Livermore National Laboratory under contract No. W-7405-ENG-48, Lawrence Berkeley National Laboratory under contract No. DE-AC03-76SF00098, and Los Alamos National Laboratory under contract No. W-7405-ENG-36.

References

1. Braun, P. and LaBaer J. (2003) High throughput protein production for functional proteomics. *Trends Biotech. Sep*; **21**(9):383-8.
2. Smith MC, Furman TC, Ingolia TD, Pidgeon C. (1988) Chelating peptide-immobilized metal ion affinity chromatography. A new concept in affinity chromatography for recombinant proteins. *J. Biol. Chem.* **263**(15):7211-5.
3. Doyle SA, Murphy MB, Massi JM, Richardson PM. (2002) High-throughput proteomics: a flexible and efficient pipeline for protein production. *J. Proteome Res.* **1**(6):531-6.
4. Studier FW, Moffatt BA. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes *J. Mol. Biol.* **189** (1):113-30.

Fig. 1. Purified proteins imaged using A) Agilent 2100 Bioanalyzer and B) traditional SDS-PAGE. Panel A: note that the lowest two bands and the uppermost band are internal standards for sizing and quantitation. Panel B: 3-15% polyacrylamide gel stained with GelCode Blue. In both panels the predicted size of the protein is labeled at the top of each lane.

